

Linearizing layers of radial binary classifiers with movable centers

Leon Bobrowski^{1,2} · Magdalena Topczewska¹

Received: 31 March 2014 / Accepted: 16 April 2015 / Published online: 15 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Ranked layers of binary classifiers are used for the linearization of learning sets composed of multivariate feature vectors. After transformation by ranked layer, each learning set can be separated by a hyperplane from the sum of other learning sets. Ranked layers can be designed, among others, from radial binary classifiers. This work elaborates on designing ranked layers from radial binary classifiers with movable centers.

Keywords Ranked layers · Linear separability · Radial binary classifiers · Movable centers

1 Introduction

Learning sets in classification problems contain examples of objects assigned to particular categories (classes). Objects are typically represented in a standardized manner by multivariate feature vectors of the same dimension. Binary classifiers transform feature vectors into numbers equal to

one or zero. Classifiers can be designed based on learning data sets according to various pattern recognition methods [1, 2].

A layer of binary classifiers aggregates input data sets if many feature vectors are transformed into the same output vector with binary components. The aggregation is separable if and only if some of the feature vectors belonging to the same class are aggregated into a single output vector. Ranked layers allow aggregating learning sets in a linearly separable manner [3]. This means that each of the learning sets may be separated from the sum of other learning sets by a hyperplane after transformation by the ranked layer. Linearly separable aggregation plays a special role in pattern recognition methods based on neural network models. In particular, this concept is important in the *perceptron* model based on formal neurons [4].

The linear separability of learning sets is important also in *support vector machines (SVM)*, one of the most popular methods in data mining, [5, 6]. An essential part of the *SVM* methods is linear separability induction through kernel functions. Selection of the appropriate kernel functions is still an open and difficult problem in many practical applications of *support vector machines*. The ranked layers can be treated as a useful alternative for the kernel functions technique in the *SVM* context.

A family of K disjoint learning sets can always be transformed into K linearly separable sets as a result of transformation by the ranked layer of formal neurons—as proved in the paper [7]. This result was extended to the ranked layers of arbitrary binary classifiers in the work [3]. The procedure of ranked layer designing from radial binary classifiers was proposed in the work [8]. An extension of this procedure to radial binary classifiers with movable centers is discussed in this work.

This work was partially supported by the S/WI/2/2013 grant from the Białystok University of Technology and by the 4.2/St/2014 grant from the Institute of Biocybernetics and Biomedical Engineering PAS.

✉ Magdalena Topczewska
m.topczewska@pb.edu.pl

Leon Bobrowski
l.bobrowski@pb.edu.pl

¹ Faculty of Computer Science, Białystok University of Technology, Wiejska 45A Str., 15-351 Białystok, Poland

² Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

2 Separable and linearly separable learning sets

Let us assume that each object O_j ($j = 1, \dots, m$) is represented in a standard manner by feature vectors $\mathbf{x}_j[n] = [x_{j1}, \dots, x_{jn}]^T$ belonging to n -dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$). Each feature vector $\mathbf{x}_j[n]$ can be treated as a point of the feature space $F[n]$. Components x_{ji} of the feature vector $\mathbf{x}_j[n]$ are expected to be numerical results of n standardized examinations of a given object O_j related to particular features x_i ($i = 1, \dots, n$) ($x_{ji} \in \{0, 1, \dots, R\}$ or $x_{ji} \in R$). In practice, we can often assume that the feature space $F[n]$ is equal to n -dimensional real space R^n ($F[n] = R^n$).

Let us assume that each object O_j belongs to one of K categories (classes) ω_k ($k = 1, \dots, K$). All the feature vectors $\mathbf{x}_j[n]$ that represent the objects O_j from one class ω_k can be collected as k th learning set C_k :

$$C_k = \{\mathbf{x}_j[n] : j \in J_k\} \quad (1)$$

where J_k is the set of indices j of objects O_j assigned to the k th class ω_k .

The learning set C_k contains m_k feature vectors $\mathbf{x}_j[n]$ assigned to the k th category ω_k . The assignment of feature vectors $\mathbf{x}_j[n]$ to particular categories ω_k can be seen as additional knowledge in the classification problem [1].

Definition 1 The learning sets C_k (1) are *separable* in the feature space $F[n]$, if they are disjoint in this space ($C_k \cap C_{k'} = \emptyset$, if $k \neq k'$). This means that the feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_{j'}[n]$ belonging to different learning sets C_k and $C_{k'}$ cannot be equal:

$$(k \neq k') \Rightarrow (\forall j \in J_k) \quad \text{and} \quad (\forall j' \in J_{k'}) \quad \mathbf{x}_j[n] \neq \mathbf{x}_{j'}[n] \quad (2)$$

We also take into consideration the separation of the learning sets C_k (1) by hyperplanes $H(\mathbf{w}_k[n], \theta_k)$ in the feature space $F[n]$

$$H(\mathbf{w}_k[n], \theta_k) = \{\mathbf{x}[n] : \mathbf{w}_k[n]^T \mathbf{x}[n] = \theta_k\} \quad (3)$$

where $\mathbf{w}_k[n] = [w_{k1}, \dots, w_{kn}]^T \in R^n$ is the *weight vector*, $\theta_k \in R^1$ is the *threshold*, and $\mathbf{w}_k[n]^T \mathbf{x}[n]$ is the inner product.

Definition 2 The feature vector $\mathbf{x}_j[n]$ is situated on the positive side of the hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3) if and only if $\mathbf{w}_k[n]^T \mathbf{x}_j[n] > \theta_k$. Similarly, the vector $\mathbf{x}_j[n]$ is situated on the negative side of $H(\mathbf{w}_k[n], \theta_k)$ if and only if $\mathbf{w}_k[n]^T \mathbf{x}_j[n] < \theta_k$.

Definition 3 The learning sets (1) are *linearly separable* in the n -dimensional feature space $F[n]$ if each of the sets

C_k can be fully separated from the sum of the remaining sets C_i by some hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3):

$$\begin{aligned} (\forall k \in \{1, \dots, K\}) \exists (\mathbf{w}_k[n], \theta_k) (\forall \mathbf{x}_j[n] \in C_k) \mathbf{w}_k[n]^T \mathbf{x}_j[n] > \theta_k \\ \text{and } (\forall \mathbf{x}_j[n] \in C_i, i \neq k) \mathbf{w}_k[n]^T \mathbf{x}_j[n] < \theta_k \end{aligned} \quad (4)$$

If the inequalities (4) hold, then all vectors $\mathbf{x}_j[n]$ from learning set C_k are situated on the positive side of hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3) and all vectors $\mathbf{x}_j[n]$ from the remaining sets C_i are situated on the negative side of this hyperplane.

3 Radial binary classifiers

The radial binary classifier $RC(\mathbf{w}_i[n], \rho_i)$ can be characterized by the sphere with the *center* $\mathbf{w}_i[n] = [w_{i1}, \dots, w_{in}]^T$ and radius ρ_i ($\rho_i > 0$) [1]. The *decision rule* $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ of radial binary classifier $RC(\mathbf{w}_i[n], \rho_i)$ is based on the distances $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ between point $\mathbf{x}[n]$ and the center $\mathbf{w}_i[n]$:

$$r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n]) = \begin{cases} 1 & \text{if } \delta(\mathbf{w}_i[n], \mathbf{x}[n]) \leq \rho_i \\ 0 & \text{if } \delta(\mathbf{w}_i[n], \mathbf{x}[n]) > \rho_i \end{cases} \quad (5)$$

In accordance with the decision rule $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$, the radial classifier $RC(\mathbf{w}_i[n], \rho_i)$ is activated by input vector $\mathbf{x}[n]$ ($r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n]) = 1$) if and only if the distance $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ between vector $\mathbf{x}[n]$ and the center $\mathbf{w}_i[n]$ is not greater than the radius ρ_i . The decision rule $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (5) of radial classifier $RC(\mathbf{w}_i[n], \rho_i)$ depends on the $n + 1$ parameters $\mathbf{w}_i[n] = [w_{i1}, \dots, w_{in}]^T$ and ρ_i .

We can also take into consideration the radial binary classifiers with a *complementary decision rule* $r^c(\mathbf{w}_i[n], \rho_i)$ of the following form:

$$r^c(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n]) = \begin{cases} 1 & \text{if } \delta(\mathbf{w}_i[n], \mathbf{x}[n]) \geq \rho_i \\ 0 & \text{if } \delta(\mathbf{w}_i[n], \mathbf{x}[n]) < \rho_i \end{cases} \quad (6)$$

The decision rules $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (5) or $r^c(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (6) depend on the distance function $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$. A few examples of popular distance functions $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ are given below [9]:

$$\begin{aligned} \delta_E(\mathbf{w}_i[n], \mathbf{x}[n]) &= ((\mathbf{x}[n] - \mathbf{w}_i[n])^T (\mathbf{x}[n] - \mathbf{w}_i[n]))^{\frac{1}{2}} \\ &\quad \times \text{the Euclidean dist.} \\ \delta_{L_1}(\mathbf{w}_i[n], \mathbf{x}[n]) &= \sum_{l=1, \dots, n} |w_{il} - x_l| \quad \text{the } L_1 \text{ dist.} \\ \delta_M(\mathbf{w}_i[n], \mathbf{x}[n]) &= ((\mathbf{x}[n] - \mathbf{w}_i[n])^T \Sigma^{-1} (\mathbf{x}[n] - \mathbf{w}_i[n]))^{\frac{1}{2}} \\ &\quad \times \text{the Mahalanobis dist.} \end{aligned} \quad (7)$$

where Σ is the covariance matrix $n \times n$ designed based on m feature vectors $\mathbf{x}_j[n]$.

The Euclidean distance function (7) is used to design radial classifiers.

4 Layers of radial binary classifiers

The layer composed of L radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with the decision rules $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (5) produces output vectors $\mathbf{r}[L]$ with L binary components r_i ($r_i \in \{0, 1\}$):

$$\mathbf{r}[L] = [r_1, \dots, r_L]^T = [r(\mathbf{w}_1[n], \rho_1; \mathbf{x}[n]), \dots, r(\mathbf{w}_L[n], \rho_L; \mathbf{x}[n])]^T \quad (8)$$

The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ transforms feature vectors $\mathbf{x}_j[n]$ from learning sets C_k (1) into sets R_k of the binary output vectors $\mathbf{r}_j[L]$:

$$R_k = \{\mathbf{r}_j[L] : \mathbf{x}_j[n] \in C_k \text{ (1)}\} \quad (9)$$

where

$$(\forall j \in \{1, \dots, m\}) \mathbf{r}_j[L] = [r(\mathbf{w}_1[n], \rho_1; \mathbf{x}_j[n]), \dots, r(\mathbf{w}_L[n], \rho_L; \mathbf{x}_j[n])]^T \quad (10)$$

Definition 4 The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (5) is *separable*, if it preserves separability (2) of learning sets C_k (1) once they are transformed into sets R_k (9). This means that the below implication is preserved after transformation (10) by the layer:

$$(k \neq k') \Rightarrow (\forall j \in J_k) \text{ and } (\forall j' \in J_{k'}) \mathbf{r}_j[L] \neq \mathbf{r}_{j'}[L] \quad (11)$$

Definition 5 The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is *linearly separable*, if the separable learning sets C_k (2) become linearly separable sets R_k (9) after transformation (10) by this layer:

$$(\forall k \in \{1, \dots, K\}) (\exists \mathbf{v}_k[L], \theta_k) (\forall \mathbf{r}_j[L] \in R_k) \mathbf{v}_k[L]^T \mathbf{r}_j[L] > \theta_k \text{ and } (\forall \mathbf{r}_j[L] \in R_i, i \neq k) \mathbf{v}_k[L]^T \mathbf{r}_j[L] < \theta_k \quad (12)$$

Each linearly separable (12) layer of binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is also a separable layer (10).

5 Designing ranked layers of radial binary classifiers

The procedure of ranked layer designing from binary radial classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (5) was proposed and described in paper [8]. This procedure was based on the examination of homogeneity of open Euclidean balls $B_j(\mathbf{x}_j[n], \rho_j)$ centered at particular feature vectors $\mathbf{x}_j[n]$:

$$(\forall j = 1, \dots, m) B_j(\mathbf{x}_j[n], \rho_j) = \{x[n] : (\mathbf{x}[n] - \mathbf{x}_j[n])^T (\mathbf{x}[n] - \mathbf{x}_j[n]) < \rho_j^2\} \quad (13)$$

Definition 6 The open Euclidean ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) is *homogeneous* in respect to learning sets C_k (1) if it contains such feature vectors $\mathbf{x}_j[n]$ that belong to only one of these sets. The ball $B_j(\mathbf{x}_j[n], \rho_j)$ is *not homogeneous* if it contains feature vectors $\mathbf{x}_j[n]$ from more than one learning set C_k (1).

In order to achieve a high *generalization power* of the ranked layer, the below designing postulate concerning the homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) was introduced [8]:

Designing postulate 1: The ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) should contain a large number of feature vectors $\mathbf{x}_j[n]$ belonging to only one of the learning sets C_k (1). (14)

In accordance with the above postulate, the largest radius ρ_i was selected for each ball $B_j(\mathbf{x}_j[n], \rho_j)$ while ensuring the homogeneity condition.

$$\rho_j = \max\{\rho : \text{the ball } B_j(\mathbf{x}_j[n], \rho_j) \text{ (13) is homogeneous}\} \quad (15)$$

The homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) contains M_j feature vectors $\mathbf{x}_j[n]$ from one of the K learning sets C_k (2) ($\mathbf{x}_j[n] \in C_k$). The optimal homogeneous ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ contains feature vectors $\mathbf{x}_j[n]$ from the k^* th learning set C_{k^*} and is characterized by the maximal number M_{j^*} of feature vectors $\mathbf{x}_j[n]$ among the other homogeneous balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13):

$$(\forall j \in \{1, \dots, m\}) M_{j^*} \geq M_j \quad (16)$$

The multistage procedure of the ranked layers designed from binary radial classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (5), proposed in paper [8], is described below:

Procedure of ranked layer designing (17)

Stage 1. (Start)

- Put $l = 1$ and define sets $D_k(l) : (\exists k \in \{1, \dots, K\}) D_k(l) = C_k$ (1)

Stage 2. (Optimal homogeneous ball $B_{j^}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13))*

- Find parameters k^* , j^* and ρ_{j^*} of the reduced data set $D_{k^*}(l)$ and the optimal homogeneous ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13). The parameter k^* ($k \in \{1, \dots, K\}$) defines the index $k(l)$ of data set $D_{k^*}(l)$ reduced during the l th step: $k(l) = k^*$ (18)

The parameters j^* and ρ_{j^*} define the reducing ball $B_l(\mathbf{x}_{j(l)}[n], \rho_{j(l)})$ (13) during the l th step:

$$j(l) = j^* \quad (19)$$

and

$$\rho_{j(l)}(l) = \rho_{j^*} \quad (20)$$

Stage 3. (Reduction of the set $D_{k^}(l)$)*

- Remove feature vectors $\mathbf{x}_j[n]$ contained in the optimal ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13)

$$\begin{aligned} D_{k^*}(l+1) &= D_{k^*}(l) - \{\mathbf{x}_j[n] : \mathbf{x}_j[n] \in B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*}) \text{ (13)}\} \\ &\text{and } (\forall k \in \{1, \dots, K\} \text{ where } k \neq k^*) D_k(l+1) = D_k(l) \end{aligned} \quad (21)$$

Stage 4. (Stop criterion)

if all data sets $D_k(l+1)$ are empty, **then stop**

else increase the index l by one ($l \rightarrow l+1$) **and go to Stage 2.**

Remark 1 Each radial binary classifier $RC(\mathbf{w}_i[n], \rho_i)$ (5) added to the layer in accordance with the procedure (17) reduces (18) data set $D_{k^*}(l)$ by at least one feature vector $\mathbf{x}_{j^*}[n]$.

It can be proved on the basis of the above *Remark 1* that if the learning sets C_k (1) are separable (2), then after finite number L steps, the procedure will be stopped. The following Lemma results [8]:

Lemma 1 The number L of radial binary classifiers $RC(\mathbf{x}_{j(l)}[n], \rho_{j(l)})$ with the decision rules $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (5) in the ranked layer is no less than the number K of learning sets C_k (1) and no greater than the number m of feature vectors $\mathbf{x}_j[n]$ in these sets.

$$K \leq L \leq m \quad (22)$$

The minimal number $L = K$ of radial binary classifiers $RC(\mathbf{x}_{j(l)}[n], \rho_{j(l)})$ (5) appears in the ranked layer when whole learning sets C_k (1) are reduced (21) during successive steps l . The maximal number $L = m$ of elements appears in the ranked layer when only single elements $\mathbf{x}_j[n]$ are reduced during successive steps l .

Theorem 1 The sets R_k (9) obtained as a result of transformation (8) of separable learning sets C_k (2) by the ranked layer (17) of L radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with the decision rules $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (5) are linearly separable (12) with thresholds θ_k equal to zero:

$$\begin{aligned} (\forall k \in \{1, \dots, K\}) (\exists \mathbf{v}_k[L], \theta_k) (\forall \mathbf{r}_j[L] \in R_k) \mathbf{v}_k[L]^T \mathbf{r}_j[L] > 0 \\ \text{and } (\forall \mathbf{r}_j[L] \in R_i, i \neq k) \mathbf{v}_k[L]^T \mathbf{r}_j[L] < 0 \end{aligned} \quad (23)$$

Proof The proof is based on the choice of such vector parameters $\mathbf{v}_k[L] = [v_{k,1}, \dots, v_{k,L}]^T$ which assure fulfilling of the inequalities (20) [3]. Let us introduce for this purpose the L -dimensional vector $\mathbf{a} = [a_1, \dots, a_L]^T$ with components a_i specified below:

$$(\forall l \in \{1, \dots, L\}) \quad a_l = 1/2^l \quad (24)$$

The weight vectors $\mathbf{v}_k = [v_{k,1}, \dots, v_{k,L}]^T$ in the inequalities (23) are defined by using the parameters $k(l)$ (18)

$$(\forall l \in \{1, \dots, L\}) \text{ if } k(l) = k, \text{ then } v_{k,1} = a_l \text{ else } v_{k,1} = -a_l \quad (25)$$

It can be directly verified that all the inequalities (23) are fulfilled by the weight vectors \mathbf{v}_k with components $v_{k,1}$ specified by the rule (25). This means that sets R_k (9) are linearly separable (4) with thresholds θ_k equal to zero. \square

The arguments formulated in works [3] and [7] have been used in the above proof of Theorem 1.

The procedure of ranked layer designing (17) allows to generate a sequence of optimal homogeneous balls $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13). The procedure (17) is stopped if each feature vector $\mathbf{x}_j[n]$ is located in optimal ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$.

The postulate (14) can be treated as an example of the greedy strategy aimed at ranked layer designing with a great power of generalization. A more general designing postulate can be formulated as:

Designing postulate II :

The ranked layer should include the minimal number

L (19) of radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (5)

$$(26)$$

We can also remark that the assumptions of the procedure (17) may be less restrictive in some points. First of all, the demand that all balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13) should be homogeneous can be relaxed in some limits. Not every feature vector $\mathbf{x}_j[n]$ must be placed in an optimal ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13). A small fraction of feature vectors $\mathbf{x}_j[n]$ (1) may remain beyond the balls $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$. After such kind relaxation of the procedure (17), full linear separability (23) of sets R_k (9) is no longer guaranteed. The sets R_k (9) may become almost linearly separable [10]. Taking into account, the sets R_k (9) which may not necessarily be linearly separable (23), but only almost linearly separable, may allow achieving greater generalization power of the designed layer of binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with decision rules (5) or (6) [10].

6 Radial binary classifiers with movable centers

The *procedure of ranked layer designing* (17) involves the search (*Stage 2*) for the optimal homogeneous balls $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13). Each optimal ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ should be distinguished by a large number M_{j^*} (16) of feature vectors $\mathbf{x}_j[n]$ from one of the K learning sets C_k (1).

The search for the optimal ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13) can be based on the sequencing of feature vectors $\mathbf{x}_j[n]$ (1) according to the distances $\delta(\mathbf{x}_j[n], \mathbf{x}_{j^*}[n])$ (7) from the current central vector $\mathbf{x}_{j^*}[n]$ used in the ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (Fig. 2). The symbol $\mathbf{x}_{j(b)}[n]$ ($\mathbf{x}_{j(b)}[n] \notin C_k$) stands for the closest vector to the central vector $\mathbf{x}_{j^*}[n]$ ($\mathbf{x}_{j^*}[n] \in C_k$ (1)):

$$(\forall \mathbf{x}_j[n] \notin C_k) \delta(\mathbf{x}_{j^*}[n], \mathbf{x}_{j(b)}[n]) \leq \delta(\mathbf{x}_{j^*}[n], \mathbf{x}_j[n]) \quad (27)$$

Remark 2 The maximal homogeneous ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13) with the center in point $\mathbf{x}_{j^*}[n]$ has radius ρ_{j^*} equal to $\delta(\mathbf{x}_{j^*}[n], \mathbf{x}_{j(b)}[n])$ (27):

$$\rho_{j^*} = \delta(\mathbf{x}_{j^*}[n], \mathbf{x}_{j(b)}[n]) \quad (28)$$

Remark 3 The ball $B_{j^*}(\mathbf{x}_{j^*}[n], \rho_{j^*})$ (13) with the center in point $\mathbf{c}_j[n]$ ($\mathbf{c}_j[n] = \mathbf{x}_{j^*}[n]$) and radius ρ_{j^*} (28) contains the maximal number M_{j^*} of feature vectors $\mathbf{x}_j[n]$ among all the other homogeneous balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13) centered in this point.

In some cases, the number M_j of feature vectors $\mathbf{x}_j[n]$ contained in the homogeneous ball $B_j(\mathbf{c}_j[n], \rho_j)$ (13) can be increased by the center $\mathbf{c}_j[n]$ displacement (*movable center*), where

$$(\forall j = 1, \dots, m) B_j(\mathbf{c}_j[n], \rho_j) = \{\mathbf{x}[n] : (\mathbf{x}[n] - \mathbf{c}_j[n])^T (\mathbf{x}[n] - \mathbf{c}_j[n]) < \rho_j^2\} \quad (29)$$

We can distinguish two types of procedures for center $\mathbf{c}_j[n]$ displacements:

- i. displacements based on averaging
 - ii. radial displacements
- (30)

Both of these procedures start from homogeneous ball $B_j(\mathbf{c}_j[n], \rho_j)$ (13) with the center in point $\mathbf{c}_j[n] = \mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in C_k$ (1)) and maximal radius $\rho_j = \delta(\mathbf{c}_j[n], \mathbf{x}_{j(b)}[n])$ (28) (Fig. 2).

7 The procedure of displacements based on averaging

The homogeneous ball $B_j(\mathbf{c}_j[n], \rho_j)$ (13) with radius ρ_j (28) is enlarged at the beginning of the procedure to heterogeneous ball $B_j(\mathbf{c}_j[n], K\rho_j)$, with coefficient K greater than one:

$$B_j(\mathbf{c}_j[n], \rho_j) \rightarrow B_j(\mathbf{c}_j[n], K\rho_j), \text{ where } K > 1 \quad (31)$$

The ball $B_j(\mathbf{c}_j[n], K\rho_j)$ contains $M_k(1)$ elements $\mathbf{x}_j[n]$ of learning set C_k (1) and some elements of other learning sets $C_{k'}$. The mean vector $\mathbf{m}_k(1)$ is computed on $M_k(1)$ elements $\mathbf{x}_j[n]$ of learning set C_k (1) in ball $B_j(\mathbf{c}_j[n], K\rho_j)$:

$$\mathbf{m}_k(1) = \sum_{j \in J_k(1)} \mathbf{x}_j[n] / M_k(1) \quad (32)$$

where $J_k(1)$ is the set of indices j of elements $\mathbf{x}_j[n]$ of learning set C_k (1) in ball $B_j(\mathbf{c}_j[n], K\rho_j)$ (31).

The temporary ball $B_1(\mathbf{m}_k(1), \rho_{j(1)})$ centered in point $\mathbf{m}_k(1)$ (32) is defined as:

$$B_1(\mathbf{m}_k(1), \rho_{j(1)}) = \{\mathbf{x}[n] : (\mathbf{x}[n] - \mathbf{m}_k(1))^T (\mathbf{x}[n] - \mathbf{m}_k(1)) < \rho_{j(1)}^2\} \quad (33)$$

where

$$\rho_{j(1)}^2 = (\mathbf{x}_{j(1)}[n] - \mathbf{m}_k(1))^T (\mathbf{x}_{j(1)}[n] - \mathbf{m}_k(1)) \quad (34)$$

and $\rho_{j(1)}$ is the largest distance ρ_j (28):

$$(\forall \mathbf{x}_j[n] \in B_j(\mathbf{c}_j[n], K\rho_j)) \quad \rho_j \leq \rho_{j(1)} \quad (35)$$

The following **stop criterion** is used in procedure *i*:

the temporary ball $B_1(\mathbf{m}_k(1), \rho_{j(1)})$ is homogeneous
 then Procedure *i* is **stopped**, otherwise
 vector $\mathbf{x}_{j(1)}[n]$ is removed from ball $B_j(\mathbf{c}_j[n], K\rho_j)$ (31),
 and step (31) is repeated

(36)

We can see that the above procedure will be stopped after a finite number of steps.

Working supposition: If coefficient K in enlarged ball $B_j(\mathbf{c}_j[n], K\rho_j)$ (31) is not excessively large, then homogeneous ball $B_1(\mathbf{m}_k(1), \rho_{j(1)})$ (33) obtained at the end (36) of the procedure contains **no less** elements $\mathbf{x}_j[n]$ than the initial homogeneous ball $B_j(\mathbf{c}_j[n], \rho_j)$ (13).

For certain structures of learning sets C_k (1), the number of elements $\mathbf{x}_j[n]$ in homogeneous ball $B_1(\mathbf{m}_k(1), \rho_{j(1)})$ (33) can be significantly increased as a result of procedure *i* (36). The procedure of *Displacements based on averaging* may be particularly useful, for example, in the case of learning sets C_k (1) with more general homogeneous spaces. An example of such a structure is shown in Fig. 1.

Enlarged ball $B_j(\mathbf{c}_j[n], K\rho_j)$ (31) was used in the above procedure description. In a more general formulation, this procedure can be started from any heterogeneous subset of feature vectors $\mathbf{x}_j[n]$ (1).

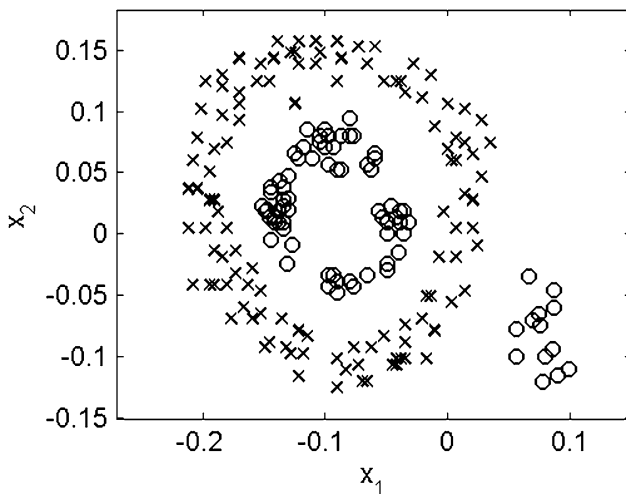


Fig. 1 Two learning sets C_1 and C_2 with more general homogeneous spaces

8 Procedures of radial displacements

This procedure can be started from any open homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) which contains M_j feature vectors $\mathbf{x}_j[n]$ from only one learning set C_k (2) ($\mathbf{x}_j[n] \in C_k$). The ball $B(\mathbf{x}_{j(a)}[n], \rho_j)$ (13) can be characterized by two feature vectors: the *central vector* $\mathbf{x}_{j(a)}[n]$ ($\mathbf{x}_{j(a)}[n] \in C_k$) and the *border vector* $\mathbf{x}_{j(b)}[n]$ ($\mathbf{x}_{j(b)}[n] \notin C_k$) with the smallest distance $\delta(\mathbf{x}_{j(a)}[n], \mathbf{x}_{j(b)}[n])$ (28) (Fig. 2).

The *central vector* $\mathbf{x}_{j(a)}[n]$ and the *border vector* $\mathbf{x}_{j(b)}[n]$ ($\mathbf{x}_{j(b)}[n] \notin C_k$) of the open homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) can be used in the following representation of this ball:

$$B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n]) = \{\mathbf{x}[n] : (\mathbf{x}[n] - \mathbf{x}_{j(a)}[n])^T (\mathbf{x}[n] - \mathbf{x}_{j(a)}[n]) < \delta^2(\mathbf{x}_{j(a)}[n], \mathbf{x}_{j(b)}[n])\} \quad (37)$$

The difference between the vectors $\mathbf{x}_{j(a)}[n]$ and $\mathbf{x}_{j(b)}[n]$ is called the *radial vector* $\mathbf{r}_{j(b),j(a)}[n]$:

$$\mathbf{r}_{j(b),j(a)}[n] = \mathbf{x}_{j(a)}[n] - \mathbf{x}_{j(b)}[n] \quad (38)$$

Vectors $\mathbf{x}_{j(a)}[n]$ and $\mathbf{x}_{j(b)}[n]$ allow to define the following *ray* $\mathbf{r}_{j(b),j(a)}(\alpha)$ in n -dimensional feature space $F[n]$ ($\mathbf{x}[n] \in F[n]$):

$$\begin{aligned} \mathbf{r}_{j(b),j(a)}(\alpha) &= \{\mathbf{x}[n] : \mathbf{x}[n] = \mathbf{x}_{j(b)}[n] + \alpha(\mathbf{x}_{j(a)}[n] - \mathbf{x}_{j(b)}[n])\} \\ &= \{\mathbf{x}[n] : \mathbf{x}[n] = \mathbf{x}_{j(b)}[n] + \alpha \mathbf{r}_{j(b),j(a)}[n]\}, \text{ where } \alpha \geq 1 \end{aligned} \quad (39)$$

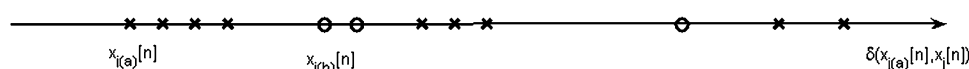


Fig. 2 An example of feature vectors $\mathbf{x}_j[n]$ (1) sequencing according to distances (7), where $\mathbf{x}_{j(a)}[n]$ ($\mathbf{x}_{j(a)}[n] \in C_k$ (1)) is the *central vector* of the homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) with four elements $\mathbf{x}_j[n]$

Radial displacement of ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37) appears when the central point $\mathbf{x}_{j(a)}[n]$ is moved along radial vector $\mathbf{r}_{j(b),j(a)}[n]$ (38). In this case, the central point $\mathbf{x}_{j(a)}[n]$ is replaced by $\mathbf{x}_\alpha[n]$:

$$\mathbf{x}_\alpha[n] = \mathbf{x}_{j(b)}[n] + \alpha \mathbf{r}_{j(b),j(a)}[n], \text{ where } \alpha \geq 1 \quad (40)$$

and the *radius* $\rho_j = \delta(\mathbf{x}_{j(a)}[n], \mathbf{x}_{j(b)}[n])$ (28) is replaced by ρ_α , where (34):

$$\rho_\alpha^2 = (\mathbf{x}_\alpha[n] - \mathbf{x}_{j(b)}[n])^T (\mathbf{x}_\alpha[n] - \mathbf{x}_{j(b)}[n]) \quad (41)$$

As a result, the ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37) is replaced by an enlarged ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$:

$$B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n]) = \{\mathbf{x}[n] : (\mathbf{x}[n] - \mathbf{x}_\alpha[n])^T (\mathbf{x}[n] - \mathbf{x}_\alpha[n]) < \rho_\alpha^2\} \quad (42)$$

Let us define the hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ tangent to the ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37) at the border point $\mathbf{x}_{j(b)}[n]$:

$$H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n]) = \{\mathbf{x}[n] : \mathbf{x}[n]^T \mathbf{r}_{j(b),j(a)}[n] = \mathbf{x}_{j(b)}[n]^T \mathbf{r}_{j(b),j(a)}[n]\} \quad (43)$$

Increasing the parameter α in the ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) can cause loss of homogeneity inherited from ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37), where $\mathbf{x}_{j(a)}[n] \in C_k$ (1). However, in some cases, homogeneity of the ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) can be preserved despite the increase of parameter α . One sufficient condition for the preservation of open ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) homogeneity during the increase of parameter α can be based on the below condition linked to tangent hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (43) and radial vector $\mathbf{r}_{j(b),j(a)}[n]$ (38):

$$\text{if } \mathbf{x}_j[n]^T \mathbf{r}_{j(b),j(a)}[n] > \mathbf{x}_{j(b)}[n]^T \mathbf{r}_{j(b),j(a)}[n], \text{ then } \mathbf{x}_j[n] \in C_k \quad (44)$$

The above condition means that each feature vector $\mathbf{x}_j[n]$ (1) situated on the positive side of tangent hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (43) belongs to the same learning set C_k as the central vector $\mathbf{x}_{j(a)}[n]$ ($\mathbf{x}_{j(a)}[n] \in C_k$) of the initial ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42). The condition (44) can be verified by computing the scalar products with radial vector $\mathbf{r}_{j(b),j(a)}[n]$ (38) and by checking the inequalities below:

$$(\forall \mathbf{x}_j[n] \notin C_k(1)) \mathbf{r}_{j(b),j(a)}[n]^T \mathbf{x}_j[n] \leq \mathbf{r}_{j(b),j(a)}[n]^T \mathbf{x}_{j(b)}[n] \quad (45)$$

and $\mathbf{x}_{j(b)}[n]$ ($\mathbf{x}_{j(b)}[n] \notin C_k$) as the *border vector* of this ball. The symbols “x” are used for $\mathbf{x}_j[n] \in C_k$ (1) and the symbols “o” are used for $\mathbf{x}_j[n] \notin C_k$

Lemma 2 If each feature vector $\mathbf{x}_j[n]$ (1) situated on the positive side (39) of tangent hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (43) belongs to the same learning set C_k as central vector $\mathbf{x}_{j(a)}[n]$ ($\mathbf{x}_{j(a)}[n] \in C_k$) of the homogeneous ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37), then the enlarged ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) is homogeneous for an arbitrarily large value of parameter α ($\alpha \geq 1$).

Lemma 2 can be proven by geometrical consideration. This Lemma can be reformulated in the manner below by using the condition (45).

Lemma 3 If each feature vector $\mathbf{x}_j[n]$ (1), which does not belong to set C_k , fulfills the condition (45), then the enlarged ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) is homogeneous for arbitrarily large values of parameter α ($\alpha \geq 1$).

If some feature vectors $\mathbf{x}_j[n]$ (1) from other learning sets $C_{k'}$ ($k' \neq k$) (1) are situated on the positive side (39) of the tangent hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (43), then the parallel shifting of this hyperplane allows to skip such a situation. Let us consider the following shifted hyperplanes $H_\beta(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ with parameter β ($\beta \geq 0$):

$$\begin{aligned} H_\beta(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n]) &= \{\mathbf{x}[n] : \mathbf{x}[n]^T \mathbf{r}_{j(b),j(a)}[n] = \mathbf{x}_\beta[n]^T \mathbf{r}_{j(b),j(a)}[n]\} \\ &= \{\mathbf{x}[n] : \mathbf{x}[n]^T \mathbf{r}_{j(b),j(a)}[n] \\ &= \mathbf{x}_{j(b)}[n]^T \mathbf{r}_{j(b),j(a)}[n] + \beta \mathbf{r}_{j(b),j(a)}[n]^T \mathbf{r}_{j(b),j(a)}[n]\} \end{aligned} \quad (46)$$

where $\mathbf{x}_\beta[n] = \mathbf{x}_{j(b)}[n] + \beta(\mathbf{x}_{j(a)}[n] - \mathbf{x}_{j(b)}[n]) = \mathbf{x}_{j(b)}[n] + \beta \mathbf{r}_{j(b),j(a)}[n]$, and $\beta \geq 1$ (39).

Remark 4 If parameter β is greater than certain threshold β_t ($\beta_t \geq 0$), then the relation (44) is fulfilled and the enlarged ball $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42) is homogeneous for arbitrarily large values of parameter α ($\alpha \geq 1$) (lemma 2).

Enlargement of the homogeneous ball $B(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (37) is aimed at increasing the number M_j of feature vectors $\mathbf{x}_j[n]$ from the learning set C_k (1) contained in this ball. Shifting (46) of the tangent hyperplane $H(\mathbf{x}_{j(a)}[n]; \mathbf{x}_{j(b)}[n])$ (38) is also done for this purpose.

9 Strategies for designing linearizing layers

The multistage procedure (17) of ranked layer designing from binary radial classifiers allows to generate, in individual steps l , the sequence of L balls $B_l(\mathbf{x}_l[n], \rho_l)$ (13) with centers $\mathbf{x}_l[n]$ and radiuses ρ_l as follows:

$$B_1(\mathbf{x}_1[n], \rho_1), B_2(\mathbf{x}_2[n], \rho_2), \dots, B_L(\mathbf{x}_L[n], \rho_L). \quad (47)$$

The balls $B_l(\mathbf{x}_l[n], \rho_l)$ are designed based on the following sequence of data sets $D_k(l)$ (17) reduced in subsequent steps l , where $D_k(1) = C_k$ (1) for $l = 1$:

$$(\forall k \in \{1, \dots, K\}) D_k(1) \supset D_k(2) \supset \dots \supset D_k(L) \quad (48)$$

Remark 5 Only one data set $D_{k(l)}(l)$ is reduced during each step l in accordance with the ranked procedure (17):

$$(\forall l = 1, \dots, L) (\forall k \neq k(l)) D_k(l) = D_k(l-1) \text{ and } D_{k(l)}(l) = D_{k(l)}(l-1)/R_{k(l)}(l-1) \quad (49)$$

where $R_{k(l)}(l-1)$ is the non-empty set of such feature vectors $\mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in D_{k(l)}(l-1)$), which are reduced during the step $l-1$.

In accordance with *Designing postulate I* (14), the set $R_{k(l)}(l-1)$ (49) should be the greatest. This means that in the context of the radial binary classifiers, the optimal ball $B_l(\mathbf{x}_l[n], \rho_l)$ (13) with center $\mathbf{x}_l[n]$ and radius ρ_l should contain the greatest number of elements $\mathbf{x}_j[n]$ of the reduced learning set $C_{k(l)}(l-1)$. The postulate (14) is an example of the *greedy* strategy aimed at designing a ranked layer with a great power of generalization. The procedure (17) of ranked layer designing includes *Designing postulate I* (14) within *Stage 2*. *Designing postulate II* (26) is somewhat more general than *Designing postulate I* (14). Postulate II (26) can lead beyond the greedy strategy, but so far there has been a lack of efficient computational procedures.

Both the procedures of *displacements based on averaging* and *radial displacements* of the homogeneous ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13) can be used to obtain a ranked layer with a great power of generalization. Two types of the above-mentioned procedures can be used alternatively for particular balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13). This means that for some ball $B_j(\mathbf{x}_j[n], \rho_j)$ (13), the best results will produce procedure displacements based on averaging, but for a different ball $B_{j'}(\mathbf{x}_{j'}[n], \rho_{j'})$ (13), better results can be achieved by using the radial displacements procedure. Typically, the best results mean the modified ball, for example the homogeneous $B(\mathbf{x}_\alpha[n]; \mathbf{x}_{j(b)}[n])$ (42), with a large number of elements $\mathbf{x}_j[n]$ of one of the sets $D_{k(l)}(l)$ (49).

A key issue remains. Which of the homogeneous balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13) should be subjected to individual procedures of displacements (30)? A variety of strategies can be proposed for the selection of one or more homogeneous balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13) and the appropriate technique to modify these balls. However, this issue requires further study.

10 Experimental results

To demonstrate the particular steps of ranked layer designing, the results of four experiments are presented. The first experiment was performed on artificial data sets with normal distributions. In the second experiment, data sets with a ring structure were used. The third experiment was carried out on

the well-known and well-understood *Iris* data set [1], and finally, three data sets from the UCI repository were chosen.

10.1 Experiment 1

In the first experiment, the procedure of radial displacements (25) of the homogeneous balls $B_j(\mathbf{x}_j[n], \rho_j)$ (13) was used. This procedure was applied to the learning sets generated in accordance with the normal model [9]. Objects belonging to two categories were randomly generated from populations with normal distributions with mean vectors $\mu_1 = [0, 0]$, $\mu_2 = [3, 1]$, respectively, and the same covariance matrices $\Sigma_1 = \Sigma_2$, where the variance of the first class equaled 2.4 and for the second 2.0, and the correlation coefficient was at level 0.9.

The results are shown in Fig. 3. The initial B_1 ball is centered at the $[-0.86, 0.64]$ point, and the initial radius is 2.54. Sixty-six objects can be correctly classified in the first step. Using the procedure with movable centers, the center is shifted to the $\mathbf{c}_1 = [-1.06, 0.69]$ point. The length of the final first radius equals $\rho_1 = 2.73$. The number of classified objects, belonging to category 1, is 71, and thus, the displacement of the center increases this number by five.

In the second step, the initial B_2 ball is centered at the $[4.11, 0.26]$ point, with a radius of 2.14. The final center in the second step is $\mathbf{c}_2 = [23.71, -21.87]$, and the radius is enlarged to $\rho_2 = 31.27$. The number of correctly classified objects in category 2 increases from 61 to 99.

The center of the final B_3 ball is $\mathbf{c}_3 = [102.33, 148.19]$, while the initial center is situated in $[2.33, 2.46]$. Using the movable centers procedure, the radius of the final B_3 ball increases from 5.48 to $\rho_3 = 180.47$. Both at the first and at the second setting, there are 17 correctly classified category 1 objects.

In the fourth step, the initial center $[-2.01, -2.67]$ is displaced to the $\mathbf{c}_4 = [-102.01, -52.14]$ point. The initial radius equaled 1.3833 and was enlarged to $\rho_4 = 112.699$. Twelve category 1 observations are correctly classified.

In the last step, the remaining one object from the second category is classified using the B_5 ball with $\mathbf{c}_5 = [-0.77, -2.06]$ center and radius 1.

10.2 Experiment 2

The procedure of displacements based on averaging was used in the second experiment. One hundred and eighty-one objects of two categories: 66 in the inner ring and 115 in the outer ring are shown in Fig. 4a.

In the first step, the center of the homogeneous ball B_1 is located in the $\mathbf{c}_1 = [-0.09, 0.05]$ point with the radius equal to $\rho_1 = 0.06$. Forty-two inner ring observations can

be correctly classified using this classifier. In the second step, the homogeneous ball $B_1(\mathbf{c}_1[n], \rho_1)$ (13) is enlarged to heterogeneous ball $B_1(\mathbf{c}_1[n], K\rho_1)$, with coefficient K greater than one. $K = 2$ is assumed. Inside the ball, there are 66 inner ring objects and 74 second category objects. By averaging the featured objects inside the ball, displacement of the center is performed. The new center is moved to the $\mathbf{c}_2 = [-0.09, 0.03]$ point. The center correction is analogical to this in the k -means method. In the last step, the radius is decreased to $\rho_2 = 0.08$. Finally, all 66 objects forming the inner ring are correctly classified. The remaining 115 outer ring objects are correctly classified using radial classifier $B_3(\mathbf{c}_3[n], \rho_3)$ with the center $\mathbf{c}_3 = [-0.10, 0.02]$ and radius $\rho_2 = 0.15$.

10.3 Experiment 3

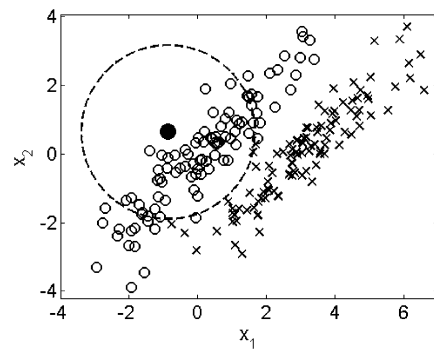
In the third experiment, the *Iris* data set was chosen. It is the well-known and well-understood problem of three species of irises, where each of 150 flowers is described by four attributes and belongs to one of three classes. For calculations, the procedure of radial binary classifiers with movable centers designing was applied.

The results are presented in Table 1. Five steps were needed to classify the objects belonging to three classes. In the first step, the whole category *Iris setosa* was perfectly classified by B_1 ball with the $\mathbf{c}_1 = [5.1, 103.5, -158.6, -89.8]$ center and the enlarged $\rho_1 = 211.14$ radius. In the second step, 48 objects belonging to the *Iris versicolor* category were classified by the B_2 ball with the $\mathbf{c}_2 = [9.7, 3.1, -1.5, -2.5]$ center and the $\rho_2 = 8.39$ radius. In the next two steps, 44 and 6 objects belonging to the *Iris virginica* category were classified by B_3 and B_4 balls ($\mathbf{c}_3 = [7.5, 3.7, 6.4, 2.7]$, $\rho_3 = 2.46$ and $\mathbf{c}_4 = [-95.1, -67.5, -25.5, -8.3]$, $\rho_4 = 127.22$). In the last step, two remaining objects belonging to the *Iris versicolor* category were correctly classified by B_5 ball ($\mathbf{c}_5 = [6.0, 2.7, 5.1, 1.6]$ and $\rho_5 = 0.63$).

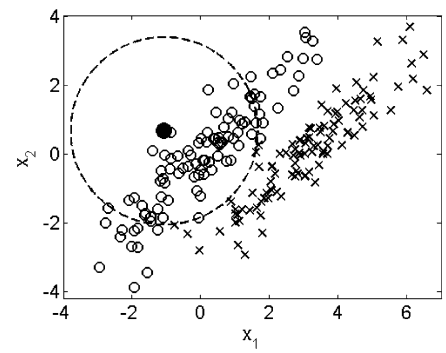
10.4 Experiment 4

In the last experiment, data sets from the UCI repository were chosen. The first data set (*Yeast*) contains data of protein localization sites in yeast bacteria, based on several bio-statistical tests. The number of objects is 1484, and each object is described by eight numerical attributes and the class label (ten classes). The objective of the second data set (*E. coli*) is similar—to predict the cellular localization sites of proteins. The data set contains 336 instances described by seven numerical attributes and the class. There are eight classes. The third chosen data set is *BreastTissue*. It presents electrical impedance measurements in samples of freshly excised tissue from the breast.

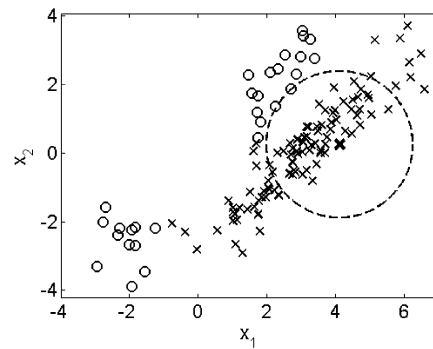
Fig. 3 Results for experiment 1—the initial and final balls of each step



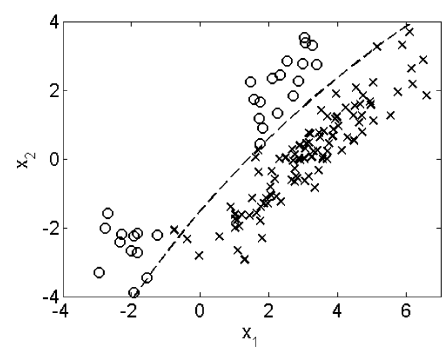
(a) Step 1 – the initial ball



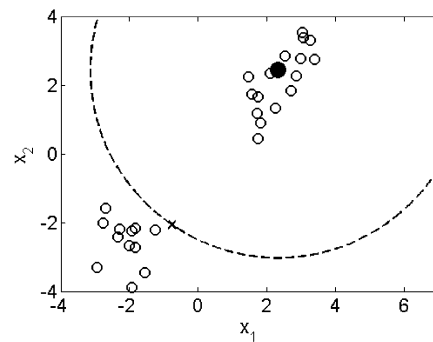
(b) Step 1 – the final ball



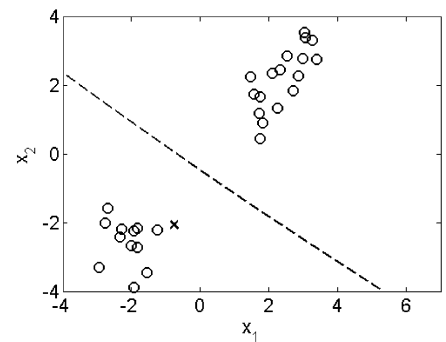
(c) Step 2 – the initial ball



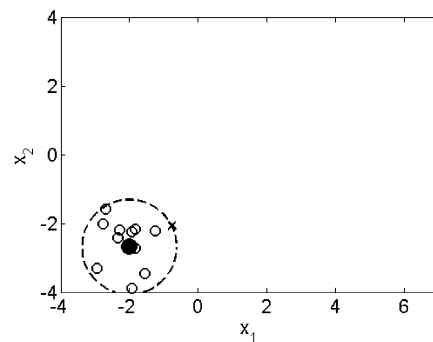
(d) Step 2 – the final ball



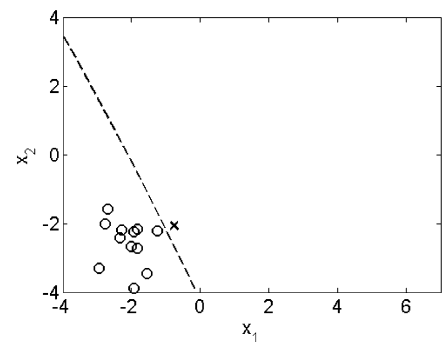
(e) Step 3 – the initial ball



(f) Step 3 – the final ball



(g) Step 4 – the initial ball



(h) Step 4 – the final ball

Fig. 4 Example results of the procedure of displacements based on averaging

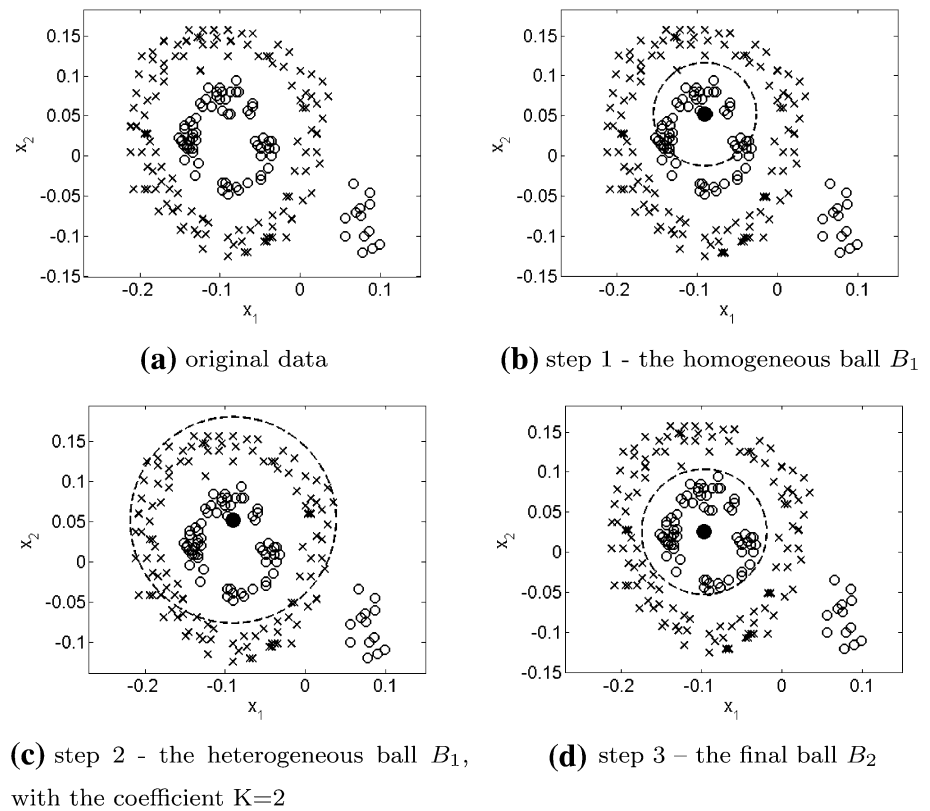


Table 1 Results for the Iris data set (\mathbf{c}_i center of B_i ball, ρ_i radius of B_i ball, m_i number of classified objects by the B_i ball)

Step i	Ball center \mathbf{c}_i	Radius ρ_i	m_i	Category
1	(5.1, 103.5, -158.6, -89.8)	211.136	50	Iris setosa
2	(9.7, 3.1, -1.5, -2.5)	8.398	48	Iris versicolor
3	(7.5, 3.7, 6.4, 2.7)	2.460	44	Iris virginica
4	(-95.1, -67.5, -25.5, -8.3)	127.215	6	Iris virginica
5	(6.0, 2.7, 5.1, 1.6)	0.625	2	Iris versicolor

One hundred and six instances, nine numerical attributes and the class are available.

The results for ranked layers of radial binary classifiers and for modification with movable centers were compared to the results of the support vector machines with the RBF kernel approach (Table 2). For the SVM method, the parameters were fixed as $C = 1.0$, $\varepsilon = 1e - 12$, $\gamma = 0.1$ or 0.5 . The classification tests were performed using the ten-fold cross-validation. To unify the results, the same shares assigned to folds were used in our own implementation and in the Weka System.

In the case of the Yeast data set, the accuracy for the ranked layers of the radial binary classifiers method as well as the movable centers approach was 0.51. The data set is complex and the number of classes is high. Not equally distributed classes and the fact that objects from various classes are not separable caused the small accuracy. The

Table 2 Results for the chosen data sets (m number of objects, n number of attributes, K number of classes, Q_{RLRBC} accuracy for the ranked layers of radial binary approach, $Q_{\text{SVM-RBF}}$ accuracy for the SVM with RBF kernel approach)

Data set	m	n	K	Q_{RLRBC}	$Q_{\text{SVM-RBF}}$
Yeast	1484	8	10	0.51	0.56
<i>E. coli</i>	336	7	8	0.79	0.76
BreastTissue	106	9	6	0.45	0.54

best results ($Q = 0.56$) were obtained using the SVM with RBF kernel approach.

The accuracy for the ranked layers of the radial binary classifiers for the *E. coli* data set was 0.79. The movable centers approach gave slightly better results ($Q = 0.80$), and it was the highest accuracy among the compared

methods. A 0.76 accuracy was obtained for the SVM with RBF kernel approach.

For the BreastTissue data set, the accuracy for the ranked layers of the radial binary classifiers was $Q = 0.45$, while for the SVM with RBF kernel approach, it was $Q = 0.54$.

This method is new and is still being researched to improve quality and determine the scope of its applicability. In our opinion, these results are encouraging for further work to optimize the strategy of ranked layer designing.

11 Concluding remarks

The ranked layer of binary classifiers allows to transform separable learning sets into sets that are linearly separable. The problem of learning set linearization is important, for example, in the context of support vector machine (SVM) techniques [5]. The linearization of the learning sets in the SVM approach is not always done successfully through a search for appropriate kernel functions.

The procedure of ranked layer designing from formal neurons was described for the first time in paper [7]. In this approach, the ranked layer was designed using hyperplanes in the feature space. The basis exchange algorithms, which are similar to linear programming, allow one to find optimal hyperplane parameters efficiently, even in the case of large multidimensional data sets.

A computationally straightforward procedure for building ranked layers using the optimal homogeneous balls was described in work [8]. This procedure is based on exhaustive examination of homogeneous balls centered in all feature vectors contained in the learning sets.

An extension of the procedure of ranked layer designing using radial binary classifiers with movable centers has been proposed and discussed in this work. In particular, center movements based on averaging and radial

displacements of the open homogeneous balls were proposed and examined. There are still many problems with this approach, but the results achieved so far are encouraging for further research and applications.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Duda OR, Hart PE, Stork DG (2001) Pattern classification. Wiley, New York
2. Fukunaga K (1972) Introduction to Statistical Pattern Recognition. Academic Press, Waltham
3. Bobrowski L (2011) Induction of linear separability through the ranked layers of binary classifiers. Engineering application of neural networks., FIP advances in information and communication technologySpringer, Berlin, pp 69–77
4. Rosenblatt F (1962) Principles of neurodynamics. Spartan Books, Washington
5. Vapnik VN (1998) Statistical learning theory. Wiley, New York
6. Hand DJ, Mannila H (2001) Principles of data mining. MIT Press, Cambridge
7. Bobrowski L (1991) Design of piecewise linear classifiers from formal neurons by some basis exchange technique. Pattern Recognit 24(9):863–870
8. Bobrowski L, Topczewska M (2013) Separable linearization of learning sets by ranked layer of radial binary classifiers. In: Burduk R et al (eds) Proceedings of the 8th international conference on computer recognition systems CORES 2013, AISCS 226. Springer, Switzerland, pp 131–140
9. Johnson RA, Wichern DW (1991) Applied multivariate statistical analysis. Prentice-Hall Inc., Englewood Cliffs
10. Bobrowski L (2007) Almost separable data aggregation by layers of formal neurons. 13th International Conference: KDS'2007: Knowledge-dialogue-solution, Varna, June 18–24, 2007. International Journal on Information Theories and Applications, Sofia, pp 34–41